

ABSTRACT

Dubbing is a term for filling the sound on the film or video. In the recording voice process in filmmaking, there are several problems in children as dubber for their character, the examples are finding a good mood for the children to be invited to the recording process, the number of children who are limited for filling the the voice, and others cause the time of filmmaking becomes longer. Therefore, we need a system that can mimic a person's voice with a high level of similarity.

In this thesis designed a voice conversion system to model the target voice signal to be emulated by the new input signal. Target sound obtained from voice recording of children aged 5 to 15 years old and the sound of the voice imitator can come from adults, teenagers, or the elderly. Feature extraction that is used is Linear Predictive Coding (LPC) and the modeling methods is Hidden Markov Model (HMM). HMM is a statistical models that are used for applications in the field of temporal pattern recognition, one of which is the analysis of sound synthesis. In the process of making the system, parts of which is synthesized are fundamental frequency (F0) and spectral.

The results of the voice conversion, the best used LPC order is the order of 19, the best state HMM modeling is the 5th state. F0 RMSE of adult men to the children after the conversion has increased by 57.7%, while the F0 RMSE of adult women to children after the conversion has increased by 15.29%. RMSE Ceptral after conversion has increased by 33.25%. In terms of similarities, MOS testing for HMM has an average value of 2.64 and in terms of quality, testing MOS for HMM has an average value of 3.23.

Keywords: Ceptral, Dubbing, HMM, LPC, MOS, RMSE, Spectral, Voice Conversion